To be published in the *International Journal of Corpus Linguistics*.

# Short Term Diachronic Shifts in Part-of-Speech Frequencies: A Comparison of the Tagged LOB and F-LOB Corpora

Christian Mair, Marianne Hundt, Geoffrey Leech, Nicholas Smith

> "Our Western civilization, it has been said, favours an overdevelopment of the intellect at the expense of the emotions. This is why people prefer nouns to verbs." (Potter 1975: 101)

## 1. Introduction

In the present paper we do not aim at the type of language-based cultural criticism encapsulated in the motto, which is taken from Potter's *Changing English* – an interesting if somewhat impressionistic account of linguistic change in present-day English written for a popular audience. Yet the quote serves as a reminder that part-of-speech frequencies in texts are far from trivial and may indeed be revealing stylistic indicators.

LOB and F-LOB are two of several one-million-word reference corpora of standard written English which have been compiled and made available by scholars associated with ICAME. The corpora resemble each other as closely as possible in size and composition, so as to ensure comparability. LOB contains a wide array of written British texts (500 text samples of ca. 2000 words each covering a range of 15 textual genres) published in 1961, whereas F-LOB samples the same types of text 30 years later.[1] The untagged material has been the subject of a number of investigations on ongoing lexico-grammatical change in British English, which in a good many instances highlighted the desirability of also having a part-of-speech-tagged version of F-LOB to compare to the tagged LOB corpus.[2] In 1998 the English Department of Freiburg University and the Department of Linguistics and Modern English Language at Lancaster agreed to pool their resources to produce this tagged version of F-LOB, the brief of Lancaster being the automated tagging with the C8 tagset,[3] a derivative of the tagset used to tag LOB, while Freiburg, in addition to providing the material, was to be responsible for the manual post-editing of tagger output. Later further comparisons will be drawn with the American counterparts of LOB and F-LOB, the Brown and Frown Corpora.[4]

---

[1] We assume that most readers will be familiar with the basic design of these standard reference corpora. More detailed information can be gleaned from the ICAME website (http://www.hit.uib.no/icame) or obtained from Hofland/ Johansson 1982, Johansson/ Hofland 1989, Johansson *et al* 1986 (LOB) or Sand/ Siemund 1992 (F-LOB).

[2] For a survey of such work see Mair 1997b, 1998 and Hundt 1998.

[3] The C8 tagset is an enriched version of the C7 tagset used for the tagging of the British National Corpus Sampler (see Garside *et al* 1997: 257-260). It avoids two weaknesses of the C7 tagset in distinguishing auxiliary uses of *be, have* and *do* from their main verb uses, and in distinguishing relative pronouns from other *wh*-pronouns.

[4] The Brown corpus ("A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers") was compiled at Brown University, Providence, Rhode Island, under the direction of W.N. Francis and Henry Kučera from 1961 to 1964 and provided the template for subsequent corpora such as LOB and F-LOB, which documented other regional varieties or other diachronic stages of the language. Frown is the 1992 update of the Brown corpus.

As the full manual post-editing is expected to take several years, it was decided to also re-tag LOB using the C8 tagset, and to undertake automated error correction procedures so as to arrive at a close approximation to the true frequencies of the tags in LOB and F-LOB.[5] This would then make possible immediate approximate comparisons of tag frequencies in the two corpora, and thus enable us to start tackling all those problems in the investigation of which consistency of tagging across corpora is a greater priority than minimising errors within corpora through manual post-editing.

The present paper represents one such pilot study. It is intended to provide a first survey of tag frequencies in LOB and F-LOB, the chief aims being (i) to document possible diachronic developments of a general nature and (ii) to identify those areas of change or variability which merit subsequent detailed investigation on the basis of the manually post-edited output.

A number of frequently conflicting claims have been made in the literature on diachronic shifts in the frequencies of parts of speech. Therefore, the results of the present investigation are expected to contribute to the study of the following, more specific issues:

(1) Prescriptivists have expressed alarm at the prospect of English succumbing to "noun disease" (Potter 1975: 101), that is an increasing information orientation that is reflected in an increasingly nominal style. If there is a factual basis to this claim, it should show in a rise in the frequency of nouns, at least in some genres, in the more recent material from F-LOB.

(2) Descriptive work on the history of English, by contrast, explicitly or implicitly suggests a contrary direction of developments. The growing use of catenative verbs, modal idioms, semi-auxiliaries, grammaticalised or grammaticalising deverbal prepositions and conjunctions, and the like, are assumed to constitute a trend towards a greater functional prominence, and hence, greater discourse frequency of the verb. Like the prescriptivists' assumption, this claim has never been investigated empirically at a level that would meet basic corpus-linguistic standards.

(3) Using their statistical multi-dimensional model of style, Biber/ Finegan (1989, 1997) have shown that various registers of written English have ‘drifted’ over three centuries towards oral stylistic norms. This ‘colloquialization’ trend is shown principally in their Factor A (informational vs. involved), where a high frequency of nouns is strongly associated with informational, written style. Verbs, on the other, tend to be associated with a more oral style. The relative infrequency of nouns in spoken language, also confirmed by Biber *et al* (1999: 61), appears therefore to have influenced some written styles since the 17th century, and we could plausibly expect that some such trend is observable over the thirty-year gap between LOB and F-LOB.

Finally, a comparative analysis of tag frequencies serves an important project-internal function, in that it is necessary in order to validate the results of investigations based on the untagged material. Most work on the untagged material has assumed a 'null hypothesis' (i.e.

[5] We considered that the 98%-accurate outcome of the automatic tagging of the two corpora by CLAWS and the Template Tagger (see Garside *et al* 1997: 102-150) did not give a sufficiently reliable result on which to base even a preliminary comparison of the two corpora. Therefore an automated error correction procedure, undertaken after the automatic tagging of the corpora by the tagging software, provides the basis for the tag frequency comparisons across the corpora in the Tables, as well as quantitative analysis used in the body of the paper. The method has been to derive, on the basis of genres of F-LOB already hand-corrected, a *corrective co-efficient* for each tag and tag category, which can then be applied to the frequency figures of genres not yet hand-corrected. However, it is as well to bear in mind that these figures are still an approximation, albeit a close one, standing in for the results of a manual post-editing not yet completed. The figures we use here are the result of a rather complex procedure which, although arguably of considerable originality in its own right, is more of technical than linguistic interest. Its description is found in Appendix 1.

that part-of-speech frequencies have remained constant in the period of observation). Thus, Mair/ Hundt 1995, for example, have interpreted the observed increase in the frequency of progressives in the press texts of LOB and F-LOB as an increase in the importance of the grammatical category, and not as a possible consequence of the fact that verbs over-all might have become more frequent, thus leading to a proportionate increase in all verbal forms, including the progressive. Given that identifiying and counting all the simple forms in the untagged corpus would have been even more tedious and time-consuming than identifying and counting the progressives, the authors of the study could not reasonably be expected to actually verify the assumed null hypothesis. Now that both corpora have been tagged, however, the null hypothesis is relatively easy to test, and, in fact, for verbs is broadly confirmed by the results presented here (see Table 1).

In this paper we will focus particularly on differences between LOB and F-LOB in the relative frequency of nouns and verbs – the parts of speech that have figured most prominently in previous discussions, as signalled by (1) - (3) above.

## 2. Tag frequencies in LOB and F-LOB: General Survey

Table 1 gives a comparison of tag frequencies for LOB and F-LOB, looking at the corpora as a whole. We have given the frequency figures for the major parts of speech only (e.g. nouns, ignoring such secondary classes as singular nouns and plural nouns). The minimal contrasts in sample length between LOB and F-LOB are such that they can scarcely distort the results of a comparison, but in any case we have supplied normalized frequencies (per million words) so that there is no question of any such distortion. The log likelihood column indicates the degree of significance of the difference between frequencies in LOB and F-LOB: a log likelihood value of 6.6 or more is significant at $p<0.01$ (Leech/ Rayson/ Wilson 2000: 17).

Table 1: Comparison of tag frequencies in LOB and F-LOB, across major subdivisions of the corpora

| WHOLE CORPUS | LOB corpus postedited | | F-LOB corpus est. from automatic | | Difference | |
|---|---|---|---|---|---|---|
| **Wordclass** | freq | per million | freq | freq per million | as % of LOB | Log likelihood |
| **Adjective** | 75391 | 74660 | 79636 | 78907 | +5.7% | 118.6 |
| **Adverb** | 62255 | 61651 | 59666 | 59120 | -4.1% | 53.6 |
| **Article** | 112933 | 111838 | 109659 | 108656 | -2.8% | 46.4 |
| **Conjunction** | 56401 | 55854 | 56298 | 55783 | -0.1% | 0.0 |
| **Determiner** | 32357 | 32043 | 29215 | 28948 | -9.7% | 158.7 |
| **Nouns** | 253911 | 251449 | 267267 | 264821 | +5.3% | 349.7 |
| **Numeral** | 16051 | 15895 | 15493 | 15351 | -3.4% | 9.6 |
| **Preposition** | 121387 | 120210 | 118004 | 116925 | -2.7% | 46.0 |
| **Pronoun** | 58224 | 57659 | 55031 | 54527 | -5.4% | 88.3 |
| **Verb** | 179911 | 178166 | 178241 | 176610 | -0.9% | 6.9 |
| **Misc** | 40971 | 40574 | 40726 | 40353 | -0.5% | 0.6 |
| **Total** | 1009792 | 1000000 | 1009236 | 1000000 | 0.0% | 0.0 |

For the frequencies of parts of speech shown in Table 2, the LOB and F-LOB corpora are broken into major genre categories, represented by categories A-C (Press), D-H (General

Prose), J (Learned) and K-R (Fiction).[6] In the case of nouns, the comparison of these subcorpora shows a uniformly upward trend from LOB to F-LOB, but with verbs the differences are more varied, with Press and Learned showing an increase in the occurrence of verbs, and General Prose and Fiction a decrease.

Table 2: Frequency of noun and verb tags in the LOB and F-LOB corpora, showing major genre subdivisions of the corpora

| Nouns | LOB corpus postedited | | F-LOB corpus est. from automatic tagging | | Difference | |
|---|---|---|---|---|---|---|
| **Subcorpus** | freq | per million | freq | freq per million | as % of LOB | Log likelihood |
| **Press*** | 52661 | 296198 | 53248 | 298675 | +0.8% | 1.8 |
| **Gen. Prose** | 107790 | 259943 | 115392 | 278885 | +7.3% | 275.9 |
| **Learned** | 42102 | 261802 | 44391 | 276865 | +5.8% | 67.6 |
| **Fiction** | 51358 | 200212 | 54236 | 211152 | +5.5% | 74.7 |
| **TOTAL** | 253911 | 251449 | 267267 | 264821 | +5.3% | (349.7) |

| Verbs | LOB corpus postedited | | F-LOB corpus est. from automatic tagging | | Difference | |
|---|---|---|---|---|---|---|
| **Subcorpus** | freq | per million | freq | freq per million | as % of LOB | Log likelihood |
| **Press*** | 29437 | 165572 | 30556 | 171392 | +3.5% | 17.9 |
| **Gen. Prose** | 69338 | 167213 | 67184 | 162373 | -2.9% | 29.4 |
| **Learned** | 24882 | 154723 | 25550 | 159357 | +3.0% | 11.0 |
| **Fiction** | 56254 | 219298 | 54951 | 213935 | -2.4% | 17.0 |
| **TOTAL** | 179911 | 178166 | 178241 | 176610 | -0.9% | (6.9) |

* Figures for Press are based on hand-corrected figures throughout.

It should be mentioned in passing that the increased frequency of nouns in F-LOB is matched by (a) a similarly consistent increase of adjective frequency, and (b) a decrease in pronoun frequency. The close relationship between noun frequency and adjective frequency is not surprising, given that the most common function of adjectives is the modification of nouns (see Biber *et al* 1999: 66). On the other hand, the *dis*association between noun and pronoun frequency is not surprising because of the competition of pronouns and full noun phrases for the same syntactic positions of subject, object and prepositional complement (see Hudson 1994, Biber *et al* 1999: 235).

We now return to the three guiding arguments (1) to (3) regarding nouns and verbs, to see how far these arguments are supported by the findings in Tables 1 and 2, as well by other comparisons.

(1) Nominal style:
In the more recent material, the greater frequency of nouns shows up clearly not only in the total for the whole corpora, but for each subdivision of the corpus. Whether this should be taken as evidence for the prescriptivist claim that English is succumbing to "noun disease" is, however, a different question. First of all, the rise in the frequency of nouns is not really an

[6] It is as well to bear in mind that these subcorpora are of different sizes; for example, A-C = 17.6% of each corpus, J = 16% of each corpus. D-H (General Prose) is the most miscellaneous subcorpus, including religious texts, popular lore, biography, official documents, etc.; J (Learned) includes scientific and other academic writing. Our division of the original 15 text categories of LOB and F-LOB into these four subcorpora follows the precedent of Hofland and Johansson (1989: 27).

alarming one overall. More fundamentally, what the prescriptivists object to as "nominal style" is not merely the frequent use of nouns (that is, a purely statistical construct), but the perceived over-use of certain types of abstract nouns, especially those derived from verbs. An answer to the question of whether English has become more "nominal" in this sense has required further searches, for derived nouns ending in *-al, -(a)tion, -ism, -ity, -ment, -sion*. From such searches we have discovered that there is indeed an increase in abstract nouns with these suffixes. However, the extent of this increase (amounting overall to about 1.03% of all nouns in LOB) does not account for more than 20% of the noun frequency difference between LOB and F-LOB.[7]

(2) Functional prominence of the verb:
The descriptive working hypothesis, if anything, fares less well than the prescriptive one. The figures show that the verb was no more and no less prominent statistically in the nineteen nineties than in 1961. However, this does not mean that the grammaticalisation and auxiliation processes which made English ever "verbier" in the past have come to a halt. Lexical searches for the appropriate forms (e.g. *going to*, *want to*, *get*) show some expected increases in frequency, although *be going to* bucks the trend, showing no increase between LOB and F-LOB.[8] It could be that the overall impression of stability is merely a reflection of the fact that these diachronic shifts in frequency are drowned out by much greater synchronic "noise" generated by variation based on genre and text-type (on which see 3 below).

(3) Stylistic drift toward oral speech norms:
The colloquialization thesis would predict not only a decrease of nouns between LOB and F-LOB, but an increase of verbs. Biber et al (1999: 65) show that across a range of present-day English registers, a considerably higher frequency of verbs is found in conversation than in

[7] The compared frequencies of these noun-forming suffixes in LOB and F-LOB are as follows:

| | LOB | F-LOB | | LOB | F-LOB |
|---|---|---|---|---|---|
| *-al* | 361 | 402 | *-ity* | 3022 | 4016 |
| *-(a)tion* | 10400 | 11146 | *-ment* | 3694 | 4210 |
| *-ism* | 483 | 819 | *-sion* | 1993 | 1985 |

Total for the six suffixes: LOB – 19953; --- F-LOB – 22576; an increase of 13.16%.

This is not a complete list of abstract-noun forming affixes, but is indicative of a general trend towards greater nominalization in F-LOB. The counts for the *–al* suffix are restricted to the deverbal nominalizing suffix, as in *proposal, refusal, betrayal.* For the other suffixes, particularly *–(a)tion*, the category is less precisely drawn, and includes, for example, words such as *nation* and *station*.

[8] For *going to* in LOB, F-LOB, Brown and Frown, compare Mair 1997a. The relevant figures for *want, get* and *going to* are given in the table below. The * character presents the wildcard; results were manually post-edited to exclude over-collected forms such as *helpful* or nominal uses of *help*. On *help*, probably the most controversial candidate for grammaticalisation among the verbs mentioned, see Mair (1995).

Below is a frequency table of selected search arguments in four corpora:

| | Brown | Frown | LOB | F-LOB |
|---|---|---|---|---|
| *want** | 579 | 909 | 681 | 777 |
| *get*/got/gotten* | 1355 | 1664 | 1408 | 1352 |
| *help** | 536 | 561 | 459 | 588 |
| *going to* | 216 | 332 | 254 | 246 |
| Total | 2686 | 3466 | 2802 | 2963 |

The comparison between Brown and Frown is instructive, in showing that the upward trend consonant with grammaticalisation is much clearer in AmE. In BrE, although the overall total is upward, the trend is more variable and fluctuating, as is shown particularly in the surprisingly lower figure for *going to* in F-LOB than in LOB.

informative writing. The results of the comparison of LOB and F-LOB frequencies, however, confirm neither part of this story. As Table 1 shows, overall, there is near stability in the frequency of verbs, but an increase close to 5% in nouns (and adjectives). This c.5% difference may appear small, but over a 30 year period it is not inconsiderable - in fact the log likelihood value in Table 1 suggests that it is highly significant.

In Table 2, the uniform increase of nouns across subcorpora shows consistency, strengthening the conviction that this is a reliable finding. However, overall, the fact that the increased frequency of nouns is not counterbalanced by a corresponding decrease in verb frequency does not fit in with the stereotypical polarity between 'nominal' and 'verbal' styles. The style of written English appears to have become more 'nominal', without becoming noticeably less 'verbal'.

(4) The null-hypothesis ("word-class frequencies remain constant")
The null-hypothesis is the only point under investigation for which previous uncertainty has been cleared up once and for all. For searches on the whole corpora, the null-hypothesis (allowing for genre adjustments) will approximately hold for verbs, while counts involving nouns and adjectives in the untagged material need to be normalised so as to offset the rise in their over-all frequency in F-LOB. Studies of parts of the corpora may be subject to greater problems, but – whatever the fluctuation in part-of-speech frequencies may be in a given sub-corpus – it is at least possible to measure the distorting influence it has on the raw frequencies obtained from searches of the untagged material. To give an example, Mair/ Hundt's (1995) study of progressives, which - based on the untagged material - noted significant increases in the press sections A-C, is not invalidated by the new findings in the tagged corpora. However, the fact that the frequency of verbs as a class increased by 7.3 % in A (and, less dramatically, by 0.7 % each in B and C) certainly accounts for part of the increase observed in the untagged material, which is thus less dramatic than it appeared in the original paper.[9]

## 3. Frequency Changes among Subcategories and Combinations of Nouns

Leaving aside discussion of other word classes, we may at this stage look more closely at the noun category from yet a further viewpoint: let us consider the frequency of different subcategories of nouns, to find out if the noun increase between LOB and F-LOB is concentrated in one subcategory rather than another:

Table 3: Frequency of selected noun subcategories in the LOB and F-LOB Corpora
Singular common nouns

| | LOB corpus | | F-LOB corpus | | Difference | |
|---|---|---|---|---|---|---|
| Subcorpus | raw freq. | per million | raw freq. | per million | %age of LOB | log likelihood |
| Press | 28047 | 157754 | 28772 | 161386 | +2.3% | 7.4 |
| Gen. Prose | 65631 | 158274 | 67996 | 164335 | +3.8% | 47.2 |
| Learned | 27254 | 169473 | 27592 | 172093 | +1.5% | 3.2 |
| Fiction | 32764 | 127726 | 34278 | 133450 | +4.5% | 32.2 |
| TOTAL | 153696 | 152206 | 158638 | 157186 | +3.3% | 80.9 |

Plural common nouns

| | LOB corpus | F-LOB corpus | Difference |
|---|---|---|---|

[9] Further ongoing research by Nicholas Smith has shown an increase of 31% in the use of the present progressive in the whole of F-LOB as compared with the whole of LOB.

| Subcorpus | raw freq. | per million | raw freq. | per million | %age of LOB | log likelihood |
|---|---|---|---|---|---|---|
| Press | 9214 | 51825 | 9835 | 55166 | +6.4% | 18.6 |
| Gen. Prose | 23844 | 57501 | 26117 | 63119 | +9.8% | 108.4 |
| Learned | 9806 | 60977 | 10783 | 67256 | +10.3% | 49.4 |
| Fiction | 8037 | 31331 | 9213 | 35868 | +14.5% | 78.7 |
| TOTAL | 50901 | 50407 | 55948 | 55436 | +10.0% | 241.3 |

Proper nouns

| | LOB corpus | | F-LOB corpus | | Difference | |
|---|---|---|---|---|---|---|
| Subcorpus | raw freq. | per million | raw freq. | per million | %age of LOB | log likelihood |
| Press | 12246 | 68879 | 12413 | 69626 | +1.1% | 0.7 |
| Gen. Prose | 14432 | 34804 | 17579 | 42486 | +22.1% | 316.9 |
| Learned | 3765 | 23412 | 4551 | 28383 | +21.2% | 76.7 |
| Fiction | 9229 | 35978 | 9474 | 36885 | +2.5% | 2.9 |
| TOTAL | 39672 | 39287 | 44017 | 43614 | +11.0% | 228.1 |

The striking feature of Table 3, as of Tables 1 and 2, is the consistency of the increase in the use of nouns across different categories and subcategories. However, although all three of these important subclasses of nouns show the same increase, they do so to markedly different degrees. The most significant increase of all is that of proper nouns, which amounts to 11%. Why the texts of F-LOB contain so many more proper nouns than the texts of LOB is not one of the questions to be answered in this article, but one suggestion which may contribute to the answer is that F-LOB reflects a greater prevalence of acronyms in the 1990s, as shown in Table 4:

Table 4: Proper nouns consisting entirely of capital letters: comparison of frequency in LOB and F-LOB

| | LOB corpus | | F-LOB corpus | | Difference | |
|---|---|---|---|---|---|---|
| Subcorpus | raw freq. | per million | raw freq. | per million | %age of LOB | Log likelihood |
| Press | 775 | 4372 | 857 | 4811 | +10.0% | 3.7 |
| Gen. Prose | 391 | 946 | 1196 | 2895 | +205.9% | 428.1 |
| Learned | 98 | 617 | 615 | 3852 | +524.1% | 414.7 |
| Fiction | 166 | 648 | 188 | 731 | +12.8% | 1.3 |
| TOTAL | 1430 | 1422 | 2856 | 2833 | +99.2% | 479.7 |

Most proper nouns which are printed entirely in capitals are acronyms: words such as UNO, UNICEF, RSPCA, etc. Although these do not make up a large proportion of all proper nouns, it is worth noting a remarkable difference between their incidence in the two corpora: acronyms appear to be nearly twice as frequent in F-LOB as in LOB.

We now illustrate another way of attacking the issue of the higher frequency of nouns in F-LOB. This is to obtain counts of noun+noun sequences, to see what change if any has taken place between LOB and F-LOB. There is more than a suspicion[10] that the favoured Germanic way of forming complex lexical expressions – the combining of nouns – is making a comeback in the later 20th century, and it may be further suspected that this change is more salient in newswriting (Press) than in other categories: witness the well-known multiple-noun headlines such as:

[10] Leonard (1968), cited in Leonard (1984:4) reports that there has been a "great increase in the occurrences of noun sequences in prose fiction from 1750 to the present day."

***BT strike threat*** *over plans to chop 1,000* (F-LOB text A06)
***Flagship hospital boss*** *out* (F-LOB text A07)

To investigate this, our first tactic was to count all tags N* N*: that is, any noun (including proper nouns) followed by other noun. The results showed a vastly significant increase in the use of noun + noun sequences in F-LOB, as shown in Table 5:

Table 5: Noun + noun sequences: comparison of frequency in the LOB and F-LOB Corpora

| | LOB Corpus | | F-LOB Corpus | | Difference | |
|---|---|---|---|---|---|---|
| | raw freq. | per million | raw freq. | per million | % of LOB | Log likelihood |
| Press | 9876 | 55714 | 10874 | 61045 | +9.6% | 43.3 |
| Gen. Prose | 12938 | 31306 | 16229 | 39277 | +25.5% | 372.8 |
| Learned | 5260 | 33127 | 5961 | 37336 | +12.7% | 40.0 |
| Fiction | 4127 | 16121 | 4952 | 19261 | +19.5% | 71.6 |
| TOTAL | 32201 | 32030 | 38016 | 37711 | +17.7% | 466.3 |

Strikingly, the most dramatic increases of noun + noun sequences are not found in Press (A-C), where it could be expected, but rather in other categories, particularly General Prose. It was decided to try other variants, but surprisingly, it was not combinations ending with a proper name, but combinations ending with a common noun that showed the steepest increase of occurrence. In Table 6, we compare LOB and F-LOB in terms of sequences of noun + common noun:

Table 6 - Sequences of Noun + Common noun: comparison of the LOB and F-LOB Corpora (excluding tags NNB, NNL*, and NNA, which are invariably associated with naming expressions)

| | LOB Corpus | | F-LOB Corpus | | Difference | |
|---|---|---|---|---|---|---|
| | raw freq. | per million | raw freq. | per million | %age of LOB | Log likelihood |
| Press | 5098 | 28760 | 6376 | 35794 | +24.5% | 136.5 |
| Gen. Prose | 8756 | 21187 | 11562 | 27982 | +32.1% | 389.4 |
| Learned | 4459 | 28083 | 5235 | 32788 | +16.8% | 58.0 |
| Fiction | 2448 | 9562 | 3366 | 13092 | +36.9% | 141.7 |
| Total | 20761 | 20651 | 26539 | 26326 | +27.5% | 691.9 |

The Table shows a very marked difference – an increase of 27.5% in F-LOB above the frequency in F-LOB. Note that the Noun + Common noun rise is a feature of every text category A-R, not just the four block groupings used in this paper; whereas Noun + Proper Noun sequences rise in only 6 of the 15 text categories.[11]

## 4. Shifts in Part-of-Speech Frequencies: Diachronic and Synchronic Factors

To cast further light on tag frequency in a diachronic perspective, it is instructive to relate the observed changes to the synchronic variation manifest in a given corpus at any one time. In their exhaustive analysis of the tagged LOB corpus, Johansson/Hofland, for example, have shown tag frequencies to vary quite drastically from genre to genre (1989: I, 7-39, in particular 15). Our figures, which are based on the C8 re-tagging of LOB and therefore differ from theirs in minor ways, are as follows:

Table 7: Noun and verb frequencies in LOB (given as percentages)

[11] The rise of Noun + Common noun sequences especially was also observed in Douglas Biber's paper based on the ARCHER historical corpus, given at the Corpus Linguistics 2001 Conference held at Lancaster from 30th March to 2nd April, 2001.

| | nouns | verbs |
|---|---|---|
| Fiction | 20.0 | 21.9 |
| Nonfiction (all) | 26.9 | 16.4 |
| Nonfiction/ press (A-C) | 29.6 | 16.6 |
| Nonfiction/ science (J) | 26.2 | 15.5 |
| Total | 25.1 | 17.8 |

In the wake of Johansson/Hofland's pioneering effort there have been a number of further corpus-based studies of part-of-speech distribution - most recently Biber et al.'s (1999) *Longman Grammar of Spoken and Written English*. None of them - including Hudson's (1994) facetiously titled "About 37% of Word-Tokens are Nouns" - casts doubt on the strong tie between genre/text-type and the frequency of nouns and verbs.[12]

Stated in the most simple terms, the major result of all such research is the following: information orientation appears to promote the use of nouns, whereas narration is characterised by a higher incidence of verbs. LOB does not contain any spoken language, so that it is impossible to ascertain without further data analysis to what extent the results from the Fiction (K-R) sections, through the incorporation of fictional dialogue, represent the situation in speech.[13] However, Leech *et al* (2001: 294-5) gives comparative percentages for the frequency of nouns and verbs as in Table 8, demonstrating that the high verb-to-noun ratio shown for fiction in Table 7 is even higher in general spoken corpus material:[14]

Table 8: Noun and verb frequencies in the BNC sampler (given as percentages)

| | nouns | verbs |
|---|---|---|
| Written texts | 28.4 | 17.3 |
| Spoken transcriptions | 14.6 | 23.1 |

What does all this mean in terms of the diachronic analysis attempted in the present paper? First and foremost, the extent of the synchronic variation observed makes clear that smallish shifts in part-of-speech ratios over time must be interpreted with extreme caution. After all, what is the significance of a 5.3% increase in nouns in the corpus overall, when at any given time there is a much greater scope for variation based on genre?

Changes in tag frequencies thus do not reflect grammatical change directly. Rather, they may hold a clue to the puzzle of how grammatical innovations spread in actual usage, namely at differential speeds through different genres. To illustrate this general assumption, consider a concrete case at hand, namely the rise in verbs of 7.3 per cent observed in our reportage samples (sections A in LOB and F-LOB). This is not a direct sign of a grammatical change, but shows a style change. Reportage over the past thirty years has moved a little closer towards other genres rich in verbs – represented by fiction and conversation in our corpora. Such colloquialisation and

[12] Hudson arrives at this genre-independent constant by redefining the "noun" category and juggling with compound frequencies of nouns and pronouns/ determiners.

[13] Part-of-speech frequency profiles of conversational English, lacking for a long time, are now available in Biber et al. (1999: 65-66, and *passim*) and Leech *et al.* (2001: 294-304).

[14] The percentages for Table 8 are arrived at by aggregating all the tags beginning with N (i.e. noun tags) and all the tags beginning with V (i.e. verb tags) in Leech *et al* (2001: 294). The frequency counts from which these figures derive are based on the spoken and written parts of the British National Corpus Sampler (consisting of one million words of spoken and one million words of written material from the BNC. The spoken data is sampled from transcriptions of a wide range of both conversational and non-conversational speech material.

informalisation of news writing is a sociocultural rather than a linguistic phenomenon - and has been plausibly accounted for by critical discourse analysts, sociologists and historians (cf., e.g., Fairclough 1992). But in due course, it will no doubt have consequences for the linguistic system, because the new stylistic climate will speed up the demise of many lexical and grammatical archaisms and prevent the establishment of new lexical and grammatical markers of more formal or literary diction.

Standard English is primarily defined through its lexicon, and through its grammar. On a textual level, however, standard English is also usage, style and choice. This is, after all, the level on which we immediately recognise the standard British English of the beginning of the 20th century and distinguish it from 1960s and 1990s English, or tell British standard English apart from American standard English – long before we confirm such first intuitions through laborious counts of grammatical or lexicogrammatical variables such as the proportion of analytical and synthetic comparatives/ superlatives or the prevalence of regularised *spoiled* and *burned* against their irregular counterparts *spoilt* and *burnt*. At this level of language change – for lack of a better term one might speak of changes in grammar-in-text -, the comparison of tag frequencies will usefully complement the quantitative study of lexical frequencies and the qualitative analysis of individual examples. In addition, the study of changing stylistic fashions and genre conventions is an interdisciplinary undertaking, linking linguistics, sociology and cultural history. The investigation of corpora may thus yield insights which are useful far beyond the field of linguistics itself, and this is a prospect we need not be unhappy about at all.

**5. Conclusion**

An immediate benefit of the tagged F-LOB corpus has turned out to be a modest but necessary one. It is now possible to gauge the extent to which shifts in the part-of-speech composition of texts between 1961 and 1991 impinge on results obtained in studies based on the untagged material.

A further substantive result of some interest is the highly significant increase in the frequency of nouns and adjectives between LOB and F-LOB. Probing further into the noun category, we have observed that the increase applies to both common and proper nouns, but that it is most significant (an increase of 11%) in the case of proper nouns. At present these findings are difficult to interpret, not being accompanied by a correspondingly substantial decrease in verbs. However, they do emphatically indicate that the expectation of a drift towards a more oral style of writing is not borne out in any increase of verbs at the expense of nouns. The increasing frequency of nouns, and above all, proper nouns is a puzzling trend which invites further research.

In the mid and long term the tagged corpus is unlikely to supersede the untagged one as a resource for descriptive linguistic research. Rather, the two corpora will complement each other. There will always be interesting research questions which cannot be translated into viable search queries even in the fine-grained language of the C8 tags and thus will have to be investigated in the untagged material. On the other hand, there is obvious potential in searches for tags and, especially, tag combinations which were impossible to retrieve from the untagged corpus.

As for the theoretically most challenging question, namely how to interpret fairly modest diachronic shifts in tag frequencies when far greater discrepancies can be shown to occur in a corpus-internal synchronic analysis of genres, further conceptual groundwork is required. What is needed is no more and no less than a model of how changing stylistic conventions and changing discourse traditions ultimately lead to changes in the underlying system of grammatical choices. A period of observation spanning thirty years will never see a grammatical change run its course but only record an episode in the spread of an innovation. On the other hand, conside-

ring our lack of solid documentation and the largely anecdotal and speculative nature of most of what we "know" about grammatical change in progress or regional differences in standard English, this is no mean achievement.

Ultimately, we hope that our findings from a comparison of tag frequencies in LOB and F-LOB (and similar corpus-based work on recent and ongoing linguistic change in standard English) will make a contribution towards a new text-oriented theory of language change. For a long time, research on syntactic change has been dominated by competence-based models such as Lightfoot's (e.g. 1979, 1999), in which mismatches between the internalised grammars of parents and children, and the consequent "imperfect" acquisition of the language by the latter, were seen as the prime force in change. In recent years, however, several performance- or utterance-based models of change have been proposed (e.g. Keller 1994, Bybee, ed. 2001, Croft 2000), which are stimulating but as yet rather general in their claims. Corpus-based investigations of specific instances of change in a well-documented language such as English will, therefore, provide one important way to check whether such models are tenable and, if so, where the specific merits and demerits of the individual proposals lie.

**References**


Biber, Douglas and Edward Finegan, 1989. Drift and the evolution of English style: a history of three genres. *Language* 65.3, 487-517.

Biber, Douglas and Edward Finegan, 1997. Diachronic relations among speech-based and written registers in English. In T. Nevalainen and L. Kahlas-Tarkka (eds.), *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen,* 253-275. Helsinki: Société Philologique.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Bybee, Joan L. (ed.) 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.

Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics,* 19(1), 61-74.

Fairclough, Norman. 1992. *Discourse and Social Change*. Cambridge: Polity.

Garside, Roger, Geoffrey Leech and Geoffrey Sampson. 1987. *The Computational Analysis of English*: *A Corpus-based Approach.* London: Longman.

Garside, Roger, Geoffrey Leech and Tony McEnery. 1997. *Corpus Annotation: Linguistic Information from Computer Text Corpora.* London: Longman.

Hofland, Knut and Stig Johansson. 1982. *Word Frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.

Hudson, Richard. 1994. About 37% of all word-tokens are nouns. *Language* 70, 331-339.

Hundt, Marianne. 1998. *New Zealand English Grammar: Fact or Fiction?* Amsterdam: Benjamins.

Keller, Rudi. 1994. *Sprachwandel: Von der unsichtbaren Hand in der Sprache*. 2nd ed. Tübingen: Francke. [English translation: *Language Change: The Invisible Hand in Language*. London: Routledge.]

Leonard, Rosemary. 1968. *The Types and Currency of Noun + Noun Sequences in Prose Usage 1750-1950.* Unpublished M.Phil thesis, University of London.

Leonard, Rosemary. 1984. *The Interpretation of English Noun Sequences on the Computer*. Amsterdam: North Holland.

Lightfoot, David. 1979. *Principles of Diachronic Syntax*. Cambridge: CUP.

Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution*. Malden, MA: Blackwell.

Johansson, Stig (in association with Eric Atwell, Roger Garside and Geoffrey Leech) 1986. *The Tagged LOB Corpus: Users' Manual*. Bergen: Norwegian Computing Centre for the Humanities.

Johansson, Stig, and Knut Hofland. 1989. *Frequency Analysis of English Vocabulary and Grammar*. 2 vols. Oxford: Clarendon.

Leech, Geoffrey, Paul Rayson and Andrew Wilson. 2001. *Word Frequencies in Written and Spoken English*. London: Longman.

Mair, Christian. 1995.Changing patterns of complementation, and concomitant grammaticalisation, of the verb *help* in present-day British English. In Bas Aarts and Charles F. Meyer (eds.) *The Verb in Contemporary English: Theory and Description*. 258-272. Cambridge: CUP.

Mair, Christian. 1997a. The spread of the *going-to*-future in written English. In Raymond Hickey and Stanisław Puppel (eds.) *Language History and Linguistic Modelling: A Festschrift for Jacek Fisiak*. 1537-1543. Berlin: Mouton de Gruyter.

Mair, Christian. 1997b. Parallel corpora: A real-time approach to the study of language change in progress. In Magnus Ljung (ed.) *Corpus-Based Studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)*. 195-209. Amsterdam: Rodopi.

Mair, Christian. 1998. Corpora and the study of the major varieties of English: Issues and results. In Hans Lindquist et al. (eds.) *The Major Varieties of English*. 139-157. Växjö: Växjö University Press.

Mair, Christian and Marianne Hundt. 1995. Why is the progressive becoming more frequent in English? A corpus-based investigation of language change in progress. *Zeitschrift für Anglistik und Amerikanistik* 43: 111-122.

Pacey, Michael, Steven Fligelstone and Paul Rayson. 1997. How to generalize the task of annotation. In Roger Garside, Geoffrey Leech and Anthony McEnery, *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman. 122-136.

Potter, Simeon. 1975. *Changing English*. 2nd ed. London: Deutsch.

Sand, Andrea, and Rainer Siemund. 1992. LOB – thirty years on … *ICAME Journal* 16: 119-122.

**Appendix 1: Documentation of Tagging and Comparison Procedures**
Although the frequency data on which this article is based are no more than approximate, we claim they are accurate enough to be relied on for all practical purposes.

The difficulty of comparing tag frequencies of LOB and F-LOB arises from the combination of the following three factors:

(a) Automatic tagging, using the CLAWS4 tagger and Template Tagger, can achieve only 98% accuracy.
(b) Although the LOB Corpus has been automatically tagged and hand-corrected, this was using an earlier tagset and an earlier version of CLAWS, which were not fully compatible with the tagging of the F-LOB Corpus, using the more advanced C8 tagset.
(c) Only part of the F-LOB Corpus (genres A-E) had been hand-corrected at the time of writing.
The following steps were taken to ensure that the tag frequency datasets were as accurate as possible:

**Step 1. Obtain the most accurate tag frequencies from the LOB Corpus**
The original published version of LOB, tagged with the CLAWS1 tagset (Garside *et al.* 1987) and fully post-edited (see Johansson *et al.* 1986) provides a good starting point for extracting comparable POS-tag frequencies. The tagging is internally consistent, follows a detailed guidelines manual, and is close to 100% correct. Even so, we found it necessary to modify this version of the corpus, to incorporate a more advanced and richer tagset, known as the C8 tagset. Some useful linguistic distinctions made in C8 are the following: the base forms of verbs are subdivided into infinitive and finite verb forms. Auxiliary uses of *be, have*, and *do* are distinguished from main-verb functions. The relativizer use of *that* is distinguished from the complementizer use. To apply these distinctions, it was necessary to run a set of context-sensitive matching rules over LOB, using the Template Tagger tool developed by Pacey/ Fligelstone /Rayson (1997). Remaining errors were corrected by hand.[15] The same method was used to bring the tagging of LOB into conformity with the current tagging manual for C8 tagging. From this new version of LOB, with its tagging revised to be as far as possible directly comparable with F-LOB, the tag frequencies were extracted, and normalised to frequencies per million to facilitate precise comparison of the two corpora.

**Step 2. Obtain the most accurate tag frequencies from the F-LOB Corpus**
A different method was required to obtain frequencies from the F-LOB Corpus, which had to be tagged from scratch, using CLAWS and the C8 tagset. The automatically tagged corpus (c.98% accurate) was passed from Lancaster to Freiburg, where the manual post-editing and correction was to be undertaken. As is well known, manual correction is a highly labour-intensive activity, and at the time of writing this article, genre categories A (newspaper reportage), B (newspaper editorial), C (newspaper reviews), D (religious) and E (popular lore) – 26.4% of the corpus - had been completed. These corrected text data from these categories were returned to Lancaster, where tag frequency counts were derived both from the automatically-tagged (uncorrected) version of genres A-E and the corrected version of the

[15] Most of the tag changes from the CLAWS1 tagset to the C8 tagset were achieved by simple one-for-one substitution: for example MD (= modal auxiliary verb) in CLAWS1 became VM in C8. However, an important exception to this was required in the area of tokenization: in the original LOB Corpus, most genitives (such as *'Gil''s*, *child're''s, today's*) were tokenized as single words, whereas in C8 such forms are tokenized as two words, each being given a separate tag, e.g.:

| | | | |
|---|---|---|---|
| CLAWS1: | Gill′s_NP$ | children′s_ NNS$ | today′s_NR$ |
| C8: | Gill_NP1 ′s_GE | children_NN2 ′s_GE | today_RT ′s_GE |

same genres. From this training corpus, the automatic-version frequency count of each tag was then divided by the corrected-version frequency count of the same tag, to obtain a *correction co-efficient* for that tag. Each correction co-efficient is a real number close to 1.0, which can than be multiplied by the automatic-version frequency for F-LOB to obtain a *projected frequency count* for the whole F-LOB corpus. In effect, the co-efficient gives the margin of error which, on the basis of the training corpus, has to be assumed for the whole corpus.

This procedure was based on the assumption that the automatic tagging system will produce the same proportion of erroneous taggings for each section of the corpus. To test out this assumption, we averaged the correction co-efficients over the three Press categories (A-C) and applied them to the General Prose category E. Since in this experiment, the training corpus (A-C) belonged to a different major genre type from the test corpus (E), it was hypothesised that if a constant error rate for each tag could not be relied on, this would show up in the experimental application of the technique to category E, for which 'true' (manually corrected) error rates were available. In practice, the outcome was satisfactory, in that the projected frequencies for category E contained an inconsiderable margin of error (the differences between the correction coefficients for noun, verbs and adjectives respectively were 0.005437, 0.002656, and 0.003874).

**Step 3. Compare the tag frequencies in LOB and F-LOB**

The degree of change between LOB and F-LOB was measured simply by differencing the frequency per million words of each tag across the two corpora. The same procedure was applied variously to groups of tags: e.g. since all tags beginning with N are nouns, and all tags beginning with V are verbs, frequencies of nouns and verbs respectively can be easily found by a search for all tags N* and V* (where * is a wild-card symbol). The test for significance used was the log likelihood test (preferable to chi-square – see Dunning 1993). Although the log likelihood values are seemingly significant for virtually every tag and tag group, it is probably safer to rely on highly significant LL values, rather than more marginal ones. This is because of well-known misgivings about the application of significance tests to corpus data, due to the complex non-random nature of textual choices.

## Appendix II: Tag Frequencies in LOB and F-LOB by Genres – Complete Listing

The following tables give the complete figures for the simplified survey provided in Table 2 in the text. The search for NN* yields all common nouns, in the singular and plural; searches for NN*1 and NN*2 give all common nouns in the singular and plural forms, respectively, while the search for NP* gives all forms of proper nouns.

Nouns

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 27959 | 27862 | -97 | -0.3 |
| B | 14299 | 15076 | +777 | +5.4 |
| C | 9745 | 9996 | +251 | +2.6 |
| D | 8322 | 8834 | +512 | +6.2 |
| E | 20551 | 21761 | +1210 | +5.9 |
| F | 23196 | 24313 | +1117 | +4.8 |
| G | 38805 | 42028 | +3223 | +8.3 |
| H | 16613 | 17776 | +1163 | +7.0 |
| J | 42455 | 44129 | +1674 | +3.9 |
| K | 11430 | 12075 | +645 | +5.6 |
| L | 9907 | 9483 | -424 | -4.3 |
| M | 2791 | 2906 | +115 | +4.1 |
| N | 12014 | 13196 | +1182 | +9.8 |
| P | 10944 | 11741 | +797 | +7.3 |
| R | 4176 | 4515 | +339 | +8.1 |

TOTAL 253207 265691 +12484 +4.9

Verbs

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 14898 | 15993 | +1095 | +7.3 |
| B | 9694 | 9766 | +72 | +0.7 |
| C | 5093 | 5128 | +35 | +0.7 |
| D | 6203 | 5481 | -722 | -11.6 |
| E | 13490 | 13166 | -324 | -2.4 |
| F | 15050 | 15024 | -26 | -0.2 |
| G | 25647 | 24307 | -1340 | -5.2 |
| H | 9592 | 9934 | +342 | +3.6 |
| J | 25318 | 25827 | +509 | +2.0 |
| K | 12923 | 12716 | -207 | -1.6 |
| L | 10859 | 11168 | +309 | +2.8 |
| M | 2444 | 2353 | -91 | -3.7 |
| N | 13016 | 12362 | -654 | -5.0 |
| P | 13827 | 13420 | -407 | -2.9 |
| R | 3382 | 3527 | +145 | +4.3 |

TOTAL 181436 180172 -1264 -0.7

Adjectives

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 6420 | 6165 | -255 | -4.0 |
| B | 4555 | 4616 | +61 | +1.3 |
| C | 3122 | 3329 | +207 | +6.6 |
| D | 2330 | 3000 | +670 | +28.8 |
| E | 6328 | 6542 | +214 | +3.4 |
| F | 6976 | 7652 | +676 | +9.7 |
| G | 12563 | 13167 | +604 | +4.8 |
| H | 4862 | 5033 | +171 | +3.5 |
| J | 13886 | 15137 | +1251 | +9.0 |
| K | 3526 | 3676 | +150 | +4.3 |
| L | 2871 | 2708 | -163 | -5.7 |
| M | 858 | 899 | +41 | +4.8 |
| N | 3368 | 3594 | +226 | +6.7 |
| P | 3180 | 3421 | +241 | +7.6 |
| R | 1378 | 1443 | +65 | +4.7 |

TOTAL 76223 80382 +4159 +5.5

NN*

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 20958 | 21034 | +76 | +0.4 |
| B | 12111 | 12348 | +237 | +2.0 |
| C | 7438 | 7502 | +64 | +0.9 |
| D | 7196 | 7549 | +353 | +4.9 |
| E | 18161 | 18589 | +428 | +2.4 |
| F | 19984 | 21011 | +1027 | +5.1 |
| G | 32740 | 33895 | +1155 | +3.5 |
| H | 15370 | 16493 | +1123 | +7.3 |
| J | 38604 | 39587 | +983 | +2.5 |
| K | 9851 | 10582 | +731 | +7.4 |
| L | 8093 | 7705 | -388 | -4.8 |
| M | 2325 | 2580 | +255 | +11.0 |
| N | 10065 | 10671 | +606 | +6.0 |
| P | 8568 | 9562 | +994 | +11.6 |
| R | 3689 | 3658 | -31 | -0.8 |

TOTAL 215153 222766 +7613 +3.5

NN*1

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 14533 | 14621 | +88 | +0.6 |
| B | 8311 | 8617 | +306 | +3.7 |
| C | 5449 | 5688 | +239 | +4.4 |
| D | 5496 | 5506 | +10 | +0.2 |
| E | 12788 | 12886 | +98 | +0.8 |
| F | 13823 | 14252 | +429 | +3.1 |
| G | 23560 | 24769 | +1209 | +5.1 |
| H | 10230 | 10947 | +717 | +7.0 |
| J | 27583 | 27740 | +157 | +0.6 |
| K | 7554 | 8053 | +499 | +6.6 |
| L | 6532 | 6109 | -423 | -6.5 |
| M | 1660 | 1868 | +208 | +12.5 |
| N | 8044 | 8208 | +164 | +2.0 |
| P | 6774 | 7489 | +715 | +10.6 |
| R | 2666 | 2734 | +68 | +2.6 |
| TOTAL | 155003 | 159487 | +4484 | +2.9 |

NN*2

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 4617 | 5022 | +405 | +8.8 |
| B | 3200 | 3179 | -21 | -0.7 |
| C | 1617 | 1650 | +33 | +2.0 |
| D | 1516 | 1877 | +361 | +23.8 |
| E | 4454 | 5070 | +616 | +13.8 |
| F | 5436 | 6050 | +614 | +11.3 |
| G | 8129 | 8298 | +169 | +2.1 |
| H | 4514 | 4864 | +350 | +7.8 |
| J | 9905 | 10801 | +896 | +9.0 |
| K | 2043 | 2318 | +275 | +13.5 |
| L | 1310 | 1390 | +80 | +6.1 |
| M | 554 | 641 | +87 | +15.7 |
| N | 1788 | 2253 | +465 | +26.0 |
| P | 1499 | 1809 | +310 | +20.7 |
| R | 920 | 817 | -103 | -11.2 |
| TOTAL | 51502 | 56039 | +4537 | +8.8 |

NP*

| | LOB | FLOB | DIFF(n) | DIFF(%) |
|---|---|---|---|---|
| A | 6921 | 6739 | -182 | -2.6 |
| B | 2146 | 2689 | +543 | +25.3 |
| C | 2297 | 2470 | +173 | +7.5 |
| D | 1118 | 1280 | +162 | +14.5 |
| E | 2338 | 3062 | +724 | +31.0 |
| F | 3145 | 3224 | +79 | +2.5 |
| G | 5971 | 8034 | +2063 | +34.6 |
| H | 1220 | 1250 | +30 | +2.5 |
| J | 3761 | 4362 | +601 | +16.0 |
| K | 1562 | 1482 | -80 | -5.1 |
| L | 1802 | 1766 | -36 | -2.0 |
| M | 466 | 322 | -144 | -30.9 |
| N | 1912 | 2489 | +577 | +30.2 |
| P | 2371 | 2169 | -202 | -8.5 |
| R | 481 | 853 | +372 | +77.3 |
| TOTAL | 37511 | 42191 | +4680 | +12.5 |